Forty-Sixth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 23-26, 2008

A Continuity Theory for Lossless Source Coding over Networks

WeiHsin Gu, Michelle Effros, and Mayank Bakshi
Department of Electrical Engineering
California Institute of Technology
Pasadena, California, CA 91125, USA
Email: {wgu, effros, mayank}@caltech.edu

***Abstract*— A continuity theory of lossless source coding over networks is established and its implications are investigated. In the given model, source and side-information random variables $\mathbf{X}$ and $\mathbf{Y}$ have finite alphabets, and the input sequences are drawn i.i.d. according to a generic distribution $P_{\mathbf{X},\mathbf{Y}}$ on $(\mathbf{X},\mathbf{Y})$. We consider traditional source coding, where all demands equal source random variables.**

We define a family of lossless source coding problems that includes prior example network source coding problems as special cases. We show that the lossless rate region $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is inner semi-continuous in $P_{\mathbf{X},\mathbf{Y}}$. We further show that for a special type of networks called super-source networks, where there is a super source node v^* that has access to $(\mathbf{X},\mathbf{Y})$ and any other node with access to some source random variable X_i is directly connected to v^*, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is also outer semi-continuous in $P_{\mathbf{X},\mathbf{Y}}$. Based on the continuity of super-source networks with respect to $P_{\mathbf{X},\mathbf{Y}}$, we conjecture that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is also outer semi-continuous and therefore continuous in $P_{\mathbf{X},\mathbf{Y}}$ for general networks.

I. Introduction

Characterization of rate regions for source coding over networks is a primary goal in the field of source coding theory. Given a network and a collection of sources and demands, the lossless rate region generalizes Shannon's source coding theorem [1] to describe the set of achievable rate vectors for which the error probability can be made arbitrarily close to zero as block length grows without bound. The lossless rate region is denoted by $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$, where $P_{\mathbf{X},\mathbf{Y}}$ is the source and side-information distribution, here assumed to be stationary and memoryless.

In this paper, which extends our earlier work from [2], we investigate the continuity of $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ with respect to $P_{\mathbf{X},\mathbf{Y}}$ for a broad family of source coding problems defined in Section II. While it is tempting to assume continuity of $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ in $P_{\mathbf{X},\mathbf{Y}}$ without proof, we note that this very basic property does not hold for all source coding problems. For example, [3] gives an example of a network source coding problem outside our proposed class for which $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is not continuous in $P_{\mathbf{X},\mathbf{Y}}$.

When $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is available in the form of a single-letter characterization, proving or disproving its continuity in $P_{\mathbf{X},\mathbf{Y}}$ is straight forward. For example, the rate region characterization of Slepian and Wolf [4] is continuous in $P_{\mathbf{X},\mathbf{Y}}$, as are the multicast for independent sources [5], dependent source [6], and dependent sources with side information [7]. Introduction of auxiliary random variables in single-letter characterizations of $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ (e.g., [8], [9], [10]) complicates the proof of continuity of $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ only mildly. Complete rate region characterizations for the simplest lossless functional source coding problem appear in [11]. While these rate regions are also continuous in the source and side information distribution, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is not continuous in $P_{\mathbf{X},\mathbf{Y}}$ for all functional source coding problems by [3].

Unfortunately, for most network source coding problems in the given class, single-letter characterizations of $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ and optimal coding strategies remain elusive, and proving continuity from limiting characterizations is less straight forward. A function is continuous if and only if it is both inner and outer semi-continuous (see definitions in Section III). Chen and Wagner demonstrate the inner semi-continuity of rate regions for Gaussian CEO problems in [12], [13]. We consider only finite-alphabet source and side-information random variables.

For our proposed class of problems, we show that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is inner semi-continuous in $P_{\mathbf{X},\mathbf{Y}}$. We further show that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is outer semi-continuous and hence continuous when $\mathcal{N}$ has a super-source node v^* that has access to the complete vector of sources $\mathbf{X} = (X_1, \ldots, X_s)$ and the complete vector of side-information random variables $\mathbf{Y} = (Y_1, \ldots, Y_t)$, and there is a direct link from v^* to any other source node v that has access to some X_i ($i \in \{1, \ldots, s\}$). We call such networks super-source networks.

The study of super-source networks may lead to more general continuity results. In particular, we propose a conjecture on super-source networks and show that if this conjecture holds then $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is outer semi-continuous, and therefore continuous for all networks in our proposed class.

The remainder of this paper is structured as follows. We formulate a class of network source coding problems and define continuity in Section II. Section III treats continuity of $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ with respect to $P_{\mathbf{X},\mathbf{Y}}$. In Section III-A, we show that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is inner semi-continuous for general non-functional source coding problems. We prove that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is continuous in $P_{\mathbf{X},\mathbf{Y}}$ for super-source networks in Section III-B. We describe the Vanishment Conjecture and investigate its implications in Section III-C. In Section III-D, we show that if true, the Vanishment Conjecture would imply that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is outer semi-continuous and therefore continuous in $P_{\mathbf{X},\mathbf{Y}}$.

[0]This material is based upon work partially supported by NSF Grant No. CCR-0325324 and Caltech's Lee Center for Advanced Networking.

978-1-4244-2926-4/08/$25.00 ©2008 IEEE

II. Formulation

Here we define a class of network source coding problems and their lossless rate regions. Let $\mathbf{X} = (X_1, \ldots, X_s)$ be the source random vector with finite alphabets $\prod_{i=1}^{r} \mathcal{X}_i$ and let $\mathbf{Y} = (Y_1, \ldots, Y_t)$ be the side-information random vector with finite alphabets $\prod_{j=1}^{t} \mathcal{Y}_j$. We assume without loss of generality that $|\mathcal{X}_i| = |\mathcal{Y}_j| = m$ for all $i \in \{1, \ldots, s\}$ and for all $j \in \{1, \ldots, t\}$. The sequence $(\mathbf{X}_1, \mathbf{Y}_1)$, $(\mathbf{X}_2, \mathbf{Y}_2)$, …is drawn i.i.d. according to a generic distribution $P_{\mathbf{X},\mathbf{Y}}$ of $(\mathbf{X}, \mathbf{Y})$, which describes all network inputs.

A directed network is an ordered pair $(\mathcal{V}, \mathcal{E})$ with vertex set $\mathcal{V}$ and edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. Vector $(v, v') \in \mathcal{E}$ if and only if there is a directed edge from v to v'. For each edge $e = (v, v') \in \mathcal{E}$, we call v the tail of e and v' the head of e, denoted by $v = \text{tail}(e)$ and $v' = \text{head}(e)$, respectively. The set of edges that end at vertex v is denoted by $\Gamma_I(v)$ and the set of edges that begin at v is denoted by $\Gamma_O(v)$, i.e.,

$$\begin{aligned} \Gamma_I(v) &:= \{e \in \mathcal{E} \;:\; \text{head}(e) = v\} \\ \Gamma_O(v) &:= \{e \in \mathcal{E} \;:\; \text{tail}(e) = v\}. \end{aligned}$$

Let $G = (\mathcal{V}, \mathcal{E})$ be a directed network. A non-functional network source coding problem $\mathcal{N}$ is defined as $\mathcal{N} = (G, \mathcal{S}, \mathcal{D})$. Here sets $\mathcal{S}$ and $\mathcal{D}$ describe the random variable availabilities and demands, respectively. The random variable availability set $\mathcal{S}$ is a subset of $\mathcal{V} \times \{X_1, \ldots, X_s, Y_1, \ldots, Y_t\}$ such that X_i $(i \in \{1, \ldots, s\})$ (resp. Y_j $(j \in \{1, \ldots, t\})$) is available at node $v \in \mathcal{V}$ if and only if $(v, X_i) \in \mathcal{S}$ (resp. $(v, Y_j) \in \mathcal{S}$). The demand set $\mathcal{D}$ is a subset of $\mathcal{V} \times \{X_1, \ldots, X_s\}$ such that node $v \in \mathcal{V}$ demands function $X_i \in \Theta$ (for some $i \in \{1, \ldots, s\}$) if and only if $(v, X_i) \in \mathcal{D}$. Let k denote the total number of reproduction demands, i.e., $k = |\mathcal{D}|$. For each $v \in \mathcal{V}$, sets $\mathcal{S}_v \subseteq \{X_1, \ldots, X_s, Y_1, \ldots, Y_t\}$ and $\mathcal{D}_v \subseteq \{X_1, \ldots, X_s\}$ summarize the random variable availabilities and demands, respectively, at node v, giving

$$\begin{aligned} \mathcal{S}_v &= \{X_i : (v, X_i) \in \mathcal{S}\} \cup \{Y_j : (v, Y_j) \in \mathcal{S}\} \\ \mathcal{D}_v &= \{X_i : (v, X_i) \in \mathcal{D}\}. \end{aligned}$$

We assume that for each demand pair $(v', X_i) \in \mathcal{D}$ $(i \in \{1, \ldots, s\})$, there exists a pair $(v, X_i) \in \mathcal{S}$ such that there is a path from v to v'.

As defined, all the demands are sources. This excludes "functional" network source coding problems like those studied in [3], [11], where demands can be functions of the sources rather than sources themselves.

A length-n block code $\mathcal{C} = (\mathcal{F}, \mathcal{G})$ for $\mathcal{N}$ with rate vector $\mathbf{R} = (R_e)_{e \in \mathcal{E}}$ contains a set of encoding functions $\mathcal{F} := \{f_e \mid e \in \mathcal{E}\}$ and a set of decoding functions $\mathcal{G} := \{g_{v,i} \mid (v, X_i) \in \mathcal{D}\}$. The encoding and decoding functions are defined as follows

(i) For each $e \in \mathcal{E}$,

$$\begin{aligned} f_e \;:\; &\prod_{e' \in \Gamma_I(\text{tail}(e))} \{1, 2, \ldots, 2^{nR_{e'}}\} \times \prod_{i : X_i \in \mathcal{S}_{\text{tail}(e)}} \mathcal{X}_i^n \\ &\times \prod_{j : Y_j \in \mathcal{S}_{\text{tail}(e)}} \mathcal{Y}_j^n \to \{1, 2, \ldots, 2^{nR_e}\}. \end{aligned}$$

(ii) For each $(v, X_i) \in \mathcal{D}$,

$$\begin{aligned} g_{v,i} \;:\; &\prod_{e \in \Gamma_I(v)} \{1, 2, \ldots, 2^{nR_e}\} \times \prod_{l : X_l \in \mathcal{S}_v} \mathcal{X}_l^n \\ &\times \prod_{j : Y_j \in \mathcal{S}_v} \mathcal{Y}_j^n \to \mathcal{X}_i^n. \end{aligned}$$

Let $\mathcal{M}$ denote the set of all probability distributions on $\prod_{i=1}^{s} \mathcal{X}_i \times \prod_{j=1}^{t} \mathcal{Y}_j$. For any set A, we use 2^A to denote the power set of A, i.e., the set of all subsets of A. Let $\mathbb{R}_+$ denote the set of nonnegative real numbers. Since for all $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is a subset of $\mathbb{R}_+^{|\mathcal{E}|}$, $\mathcal{R}_L(\cdot)$ can be considered as a function from $\mathcal{M}$ to $2^{\mathbb{R}_+^{|\mathcal{E}|}}$.

Before defining the continuity property, we introduce the definitions of set operations in $\mathbb{R}_+^{|\mathcal{E}|}$ and distances on $\mathcal{M}$ and $2^{\mathbb{R}_+^{|\mathcal{E}|}}$ used in this paper. We define some set operations in the following definitions.

Definition 2.1: Let A and B be subsets of the n-dimensional Euclidean space $\mathbb{R}^n$.

(a) *For any n-dimensional vector $\mathbf{v} \in \mathbb{R}^n$, define the set*

$$A + \mathbf{v} := \{\mathbf{a} + \mathbf{v} \mid \mathbf{a} \in A\}.$$

(b) *For any $\lambda, \mu \in \mathbb{R}_+ \cup \{0\}$, define the set*

$$\lambda A + \mu B := \{\lambda \mathbf{a} + \mu \mathbf{b} \mid \mathbf{a} \in A,\ \mathbf{b} \in B\}.$$

Definition 2.2: Given a positive integer n and a real number $r \in \mathbb{R}$. Define

$$\begin{aligned} \mathbf{r} &:= (r, \ldots, r) \in \mathbb{R}^n \\ \mathbf{0} &:= (0, \ldots, 0) \in \mathbb{R}^n. \end{aligned}$$

Definition 2.3: Let $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ be two n-dimensional real vectors.

(a) *We say that $\mathbf{a}$ is greater than or equal to $\mathbf{b}$, denoted by $\mathbf{a} \geq \mathbf{b}$, if and only if $a_i \geq b_i$ for all $i \in \{1, \ldots, n\}$.*

(b) *We say that $\mathbf{a}$ is greater than $\mathbf{b}$, denoted by $\mathbf{a} > \mathbf{b}$, if and only if $a_i > b_i$ for all $i \in \{1, \ldots, n\}$.*

The metrics on $\mathcal{M}$ and $2^{R_+^{|\mathcal{E}|}}$ used in this paper are as follows.

Definition 2.4: Let A and B be two subsets of the n-dimensional Euclidean space $\mathbb{R}^n$. We use $||\mathbf{x}||$ to denote the L_2-norm of $\mathbf{x} \in \mathbb{R}^n$. For any $\epsilon > 0$, sets A and B are said to be ϵ-close $(\epsilon > 0)$ if and only if

(a) *For every $\mathbf{a} \in A$, there exists $\mathbf{b}_0 \in B$ such that $||\mathbf{a} - \mathbf{b}_0|| \leq \epsilon\sqrt{n}$.*

(b) *For every $\mathbf{b} \in B$, there exists $\mathbf{a}_0 \in A$ such that $||\mathbf{b} - \mathbf{a}_0|| \leq \epsilon\sqrt{n}$.*

Note that the above definition is equivalent to the Hausdorff distance, which is used for compact subsets of $\mathbb{R}^n$. This notion of the "distance" between two subsets in $\mathbb{R}^n$ leads to the definitions of the continuity of function $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ (with respect to $P_{\mathbf{X},\mathbf{Y}}$), which is described below.

Definition 2.5: $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is continuous with respect to $P_{\mathbf{X},\mathbf{Y}}$ if and only if for any $\epsilon > 0$, there exists some $\delta > 0$ such that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ and $\mathcal{R}_L(Q_{\mathbf{X},\mathbf{Y}})$ are ϵ-close whenever $||P_{\mathbf{X},\mathbf{Y}} - Q_{\mathbf{X},\mathbf{Y}}|| < \delta$.

III. Continuity of $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ with respect to $P_{\mathbf{X},\mathbf{Y}}$

This section investigates the continuity of $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ with respect to $P_{\mathbf{X},\mathbf{Y}}$. We begin by defining inner semi-continuity and outer semi-continuity. One can show that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is continuous in $P_{\mathbf{X},\mathbf{Y}}$ if and only if both inner and outer semi-continuities hold.

Definition 3.1: Rate region $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is inner semi-continuous with respect to $P_{\mathbf{X},\mathbf{Y}}$ if and only if

$$\liminf_{l\to\infty} \mathcal{R}_L(P^{(l)}_{\mathbf{X},\mathbf{Y}}) \supseteq \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$$

for every sequence $\{P^{(l)}_{\mathbf{X},\mathbf{Y}}\} \subset \mathcal{M}$ that converges to $P_{\mathbf{X},\mathbf{Y}}$.

Definition 3.2: Rate region $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is outer semi-continuous with respect to $P_{\mathbf{X},\mathbf{Y}}$ if and only if

$$\limsup_{l\to\infty} \mathcal{R}_L(P^{(l)}_{\mathbf{X},\mathbf{Y}}) \subseteq \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$$

for every sequence $\{P^{(l)}_{\mathbf{X},\mathbf{Y}}\} \subset \mathcal{M}$ that converges to $P_{\mathbf{X},\mathbf{Y}}$.

A. Inner Semi-Continuity

For any length-n block code $\mathcal{C}$ and for every edge $e \in \mathcal{E}$, let F_e denote the random variable that represents the encoded message on e and $\widehat{X}_i^n(v')$ for all $(v', X_i) \in \mathcal{D}$ denote the reproduction of X_i at node v' by using the block code $\mathcal{C}$.

Suppose that $\mathcal{N}$ is a non-functional network source coding problem. Recall that for each $(v', X_i) \in \mathcal{D}$, there exits a pair $(v, X_i) \in \mathcal{S}$ and a path, denoted by $\mathcal{P}(v', X_i)$, which starts from v and ends at v'. For every $e \in \mathcal{E}$, let $\omega(e) := \{(v', X_i) \mid e \in \mathcal{P}(v', X_i)\}$. Define the set $\mathcal{R}^*_L(P_{\mathbf{X},\mathbf{Y}})$ the closure of the set below

$$\{\mathbf{R} = (R_e)_{e\in\mathcal{E}} \mid \ R_e \geq \frac{1}{n}(H(F_e)_P + \sum_{\omega(e)} H(X_i^n|\widehat{X}_i^n(v'))_P) \text{ for some length-}n\text{ block code } \mathcal{C}\}.$$

We next show that $\mathcal{R}^*_L(P_{\mathbf{X},\mathbf{Y}}) = \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ when $\mathcal{N}$ is non-functional. This infinite-dimensional characterization $\mathcal{R}^*_L(P_{\mathbf{X},\mathbf{Y}})$ is useful for proving inner semi-continuity.

Lemma 3.3: If $\mathcal{N}$ is non-functional, then

$$\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}}) = R^*_L(P_{\mathbf{X},\mathbf{Y}}).$$

Proof. By definition, any $\mathbf{R}$ in the interior of $\mathcal{R}^*_L(P_{\mathbf{X},\mathbf{Y}})$ is losslessly achievable. Thus, $\mathcal{R}^*_L(P_{\mathbf{X},\mathbf{Y}}) \subseteq \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$. Conversely, let $\mathbf{R} = (R_e)_{e\in\mathcal{E}} \in \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$. For any $\epsilon > 0$, choose a rate-$\mathbf{R}$ length-n block code $\mathcal{C}$ such that for all $(v', X_i) \in \mathcal{D}$,

$$\Pr\{X_i^n \neq \widehat{X}_i^n(v')\} < \frac{\epsilon}{mk}.$$

Then by Fano's inequality, for all $e \in \mathcal{E}$

$$R_e + \epsilon \geq \frac{1}{n}\left(H(F_e) + \sum_{(v', X_i)} H(X_i^n|\widehat{X}_i^n(v'))\right).$$

Thus $\mathbf{R} + \epsilon \cdot \mathbf{1} \in \mathcal{R}^*_L(P_{\mathbf{X},\mathbf{Y}})$. Since $\epsilon > 0$ is arbitrary, $\mathbf{R} \in \mathcal{R}^*_L(P_{\mathbf{X},\mathbf{Y}})$. □

We next use $\mathcal{R}^*_L(P_{\mathbf{X},\mathbf{Y}})$ to prove that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is inner semi-continuous when $\mathcal{N}$ is non-functional.

Theorem 3.4: Let $\mathcal{N}$ be a non-functional network source coding problem. Given a source distribution $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$. For every sequence $\{P^{(l)}_{\mathbf{X},\mathbf{Y}}\} \subset \mathcal{M}$ that converges to $P_{\mathbf{X},\mathbf{Y}}$,

$$\liminf_{l\to\infty} \mathcal{R}_L(P^{(l)}_{\mathbf{X},\mathbf{Y}}) \supseteq \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}}).$$

Proof. Let $\mathbf{R} = (R_e)_{e\in\mathcal{E}} \in \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$. By Lemma 3.3, for any $\epsilon > 0$, there exist an n and a length-n block code $\mathcal{C}$ such that

$$R_e + \epsilon \geq \frac{1}{n}\left(H(F_e)_P + \sum_{w(e)} H(X_i^n|\widehat{X}_i^n(v'))_P\right)$$

for any $e \in \mathcal{E}$. Since $\lim_{l\to\infty} P^{(l)}_{\mathbf{X},\mathbf{Y}} = P_{\mathbf{X},\mathbf{Y}}$ and n is fixed,

$$\lim_{l\to\infty} \frac{1}{n} H(F_e)_{P^{(l)}} = \frac{1}{n} H(F_e)_P$$

for any $e \in \mathcal{E}$, and $\forall (v', X_i) \in \mathcal{D}$,

$$\lim_{l\to\infty} \sum_{\omega(e)} H(X_i^n|\widehat{X}_i^n(v'))_{P^{(l)}} = \sum_{\omega(e)} H(X_i^n|\widehat{X}_i^n(v'))_P.$$

Hence there exists l' such that for all $l \geq l'$,

$$R_e + 2\epsilon \geq \frac{1}{n}\left(H(F_e)_{P^{(l)}} + \sum_{\omega(e)} H(X_i^n|\widehat{X}_i^n(v'))_{P^{(l)}}\right)$$

for any $e \in \mathcal{E}$. By Lemma 3.3,

$$\mathbf{R} + 2\epsilon \cdot \mathbf{1} \in \bigcap_{l\geq l'} \mathcal{R}_L(P^{(l)}_{\mathbf{X},\mathbf{Y}}).$$

Therefore, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}}) \subseteq \liminf_{l\to\infty} \mathcal{R}_L(P^{(l)}_{\mathbf{X},\mathbf{Y}})$. □

B. Super-Source Networks

We call network source coding problem $\mathcal{N}$ a super-source network if it has one source node v^* with access to $(\mathbf{X}, \mathbf{Y})$, and for any other node v' that has access to some X_i ($i \in \{1, \ldots, s\}$), there is a direct link from v^* to v, that is, $(v^*, v) \in \mathcal{E}$. The following definition formalizes this idea.

Definition 3.5: Network source coding problem $\mathcal{N}$ is called a super-source network if and only if $(v^*, X_i) \in \mathcal{S}$ for all $i \in \{1, \ldots, s\}$, $(v^*, Y_j) \in \mathcal{S}$ for all $j \in \{1, \ldots, t\}$, and $(v^*, v) \in \mathcal{E}$ for all $v \in \mathcal{V}\backslash\{v^*\}$ such that $(v, X_i) \in \mathcal{S}$ for some $i \in \{1, \ldots, s\}$.

We show that for any network source coding problem $\mathcal{N}$ of this type, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is continuous in $P_{\mathbf{X},\mathbf{Y}}$. We begin by showing $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is continuous in $P_{\mathbf{X},\mathbf{Y}}$ for single-super-source network coding problems, a subfamily of super-source networks defined blow. Then we extend this result to show that this continuity property holds for super-source network source coding problems. The following definition defines single-super-source networks.

Definition 3.6: Network source coding problem $\mathcal{N}$ is called a single-super-source network if and only if $(v^*, X_i) \in \mathcal{S}$ for all $i \in \{1, \ldots, s\}$, $(v^*, Y_j) \in \mathcal{D}$ for all $j \in \{1, \ldots, t\}$, and $(v, X_i) \notin \mathcal{S}$ for all $v \in \mathcal{V}\backslash\{v^*\}$ and $i \in \{1, \ldots, s\}$.

An example of a single-super-source networks is the "diamond network," shown in Fig. 1. Theorem 3.7 proves the continuity of $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ with respect to $P_{\mathbf{X},\mathbf{Y}}$. The proof of Theorem 3.7 replies on Lemma 1.1, which is stated and proved in Appendix A.

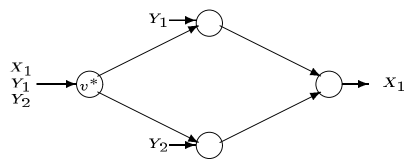


Fig. 1. The diamond network

Theorem 3.7: $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is continuous in $P_{\mathbf{X},\mathbf{Y}}$ for single-super-source network source coding problems.

Proof. To show the continuity property, it suffices to show that

$$\lim_{l\to\infty} \mathcal{R}_L(P^{(l)}_{\mathbf{X},\mathbf{Y}}) = \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$$

for any sequence $\{P^{(l)}_{\mathbf{X},\mathbf{Y}}\}$ that converges to $P_{\mathbf{X},\mathbf{Y}}$.
Consider the new network source coding problem $\mathcal{N}_1 = (G, \mathcal{S}_1, \mathcal{D}_1)$, with $2s$ source and t side-information random variables, where

$$\begin{aligned} \mathcal{S}_1 &:= \mathcal{S} \cup \{(v^*, X_i) \mid i \in \{s+1, \ldots, 2s\}\} \\ \mathcal{D}_1 &:= \mathcal{D} \cup \{(v, X_{i+s}) \mid (v, X_i) \in \mathcal{D} \\ &\qquad \text{for some } i \in \{1, \ldots, s\}\}. \end{aligned}$$

By Lemma 1.1, for any $\epsilon > 0$, choose $\delta > 0$ such that for any $Q_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$ satisfying $||Q_{\mathbf{X},\mathbf{Y}} - P_{\mathbf{X},\mathbf{Y}}|| < \delta$, there exists a random vector $(\mathbf{X}', \mathbf{Y}')$ with distribution $Q_{\mathbf{X},\mathbf{Y}}$ such that for all $i \in \{1, \ldots, s\}$ and for all $j \in \{1, \ldots, t\}$

$$\begin{aligned} &\Pr(X_i \neq X_i') < \frac{\epsilon}{d_{\max}} \\ &\Pr(Y_j \neq Y_j') < \min\left\{\tau, \frac{\epsilon}{d_{\max}}\right\}, \end{aligned} \tag{1}$$

where τ is chosen so that

$$H(1-\tau, \frac{\tau}{m-1}, \ldots, \frac{\tau}{m-1}) < \epsilon. \tag{2}$$

Let $P_{\mathbf{X},\mathbf{X}',\mathbf{Y},\mathbf{Y}'}$ be the joint distribution for $(\mathbf{X}, \mathbf{X}', \mathbf{Y}, \mathbf{Y}')$.
Let $\mathcal{R}_{L,1}(P_{\mathbf{X},\mathbf{X}',\mathbf{Y},\mathbf{Y}'})$ denote the lossless rate region for the network $\mathcal{N}_1$. Now by Lemma 1.3 (a) and (b), for $\epsilon > 0$ sufficiently small,

$$\begin{aligned} &\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}}) \subseteq \mathcal{R}_{L,1}(P_{\mathbf{X},\mathbf{X}',\mathbf{Y}}), \\ &\mathcal{R}_{L,1}(P_{\mathbf{X},\mathbf{X}',\mathbf{Y}}) + \epsilon \cdot \mathbf{1} \subseteq \mathcal{R}_L(P_{\mathbf{X}',\mathbf{Y}}). \end{aligned}$$

Hence $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}}) + \epsilon \cdot \mathbf{1} \subseteq \mathcal{R}_L(P_{\mathbf{X}',\mathbf{Y}})$.By symmetry, $\mathcal{R}_L(P_{\mathbf{X}',\mathbf{Y}}) + \epsilon \cdot \mathbf{1} \subseteq \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$.Therefore, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ and $\mathcal{R}_L(P_{\mathbf{X}',\mathbf{Y}})$ are ϵ-close.

Consider two other network source coding problems $\mathcal{N}_2 = (G, \mathcal{S}_2, \mathcal{D}_2)$, with s sources and $2t$ side-information random variables, and $\mathcal{N}_3 = (G, \mathcal{S}_3, \mathcal{D}_3)$, with $s+t$ sources and t side-information random variables. The sets $\mathcal{S}_2$, $\mathcal{S}_3$, $\mathcal{D}_2$, and $\mathcal{D}_3$ are defined as

$$\begin{aligned} \mathcal{S}_2 &:= \mathcal{S} \cup \{(v, Y_{j+t}) \mid (v, Y_j) \in \mathcal{S} \\ &\qquad \text{for some } j \in \{1, \ldots, t\}\}, \\ \mathcal{D}_2 &:= \mathcal{D}, \\ \mathcal{S}_3 &:= \{(v^*, X_i) \mid i \in \{1, \ldots, s\}\} \cup \\ &\qquad \{(v^*, Y_j) \mid j \in \{1, \ldots, 2t\}\} \cup \\ &\qquad \{(v, Y_{j+t}) \mid v \neq v^*,\ (v, Y_j) \in \mathcal{S} \\ &\qquad\quad \text{for some } j \in \{1, \ldots, t\}\}, \\ \mathcal{D}_2 &:= \mathcal{D} \cup \\ &\qquad \{(v, Y_j) \mid v \neq v^*,\ (v, Y_j) \in \mathcal{S} \\ &\qquad\quad \text{for some } j \in \{1, \ldots, t\}\}. \end{aligned}$$

Therefore by the definition of $\mathcal{D}_3$, in the network source coding problem $\mathcal{N}_3$, we require that side information Y_j should be losslessly reproduced at the nodes v' for which v' has access to Y_j in the original network source coding problem $\mathcal{N}$.
Let $\mathcal{R}_{L,2}(P_{\mathbf{X}',\mathbf{Y},\mathbf{Y}'})$ denote the lossless rate region for the network $\mathcal{N}_2$. Then since $\mathcal{N}_2$ has more side information at the nodes $v \in \mathcal{V}$ with $(v, Y_j) \in \mathcal{S}$, $\mathcal{R}_L(P_{\mathbf{X}',\mathbf{Y}'}) \subseteq \mathcal{R}_{L,2}(P_{\mathbf{X}',\mathbf{Y},\mathbf{Y}'})$.By (1) and (2), $H(Y_i|Y_i') < \epsilon$ for all $1 \leq i \leq t$. Therefore,

$$\mathcal{R}_{L,2}(P_{\mathbf{X}',\mathbf{Y},\mathbf{Y}'}) + t\epsilon \cdot \mathbf{1} \subseteq \mathcal{R}_{L,3}(P_{\mathbf{X}',\mathbf{Y},\mathbf{Y}'}).$$

By Lemma 1.3 (c), $\mathcal{R}_{L,3}(P_{\mathbf{X}',\mathbf{Y},\mathbf{Y}'}) \subseteq \mathcal{R}_L(P_{\mathbf{X}',\mathbf{Y}'})$.Hence $\mathcal{R}_L(P_{\mathbf{X}',\mathbf{Y}'})$ and $\mathcal{R}_{L,2}(P_{\mathbf{X}',\mathbf{Y},\mathbf{Y}'})$ are $(t\epsilon)$-close. By symmetry, $\mathcal{R}_L(P_{\mathbf{X}',\mathbf{Y}})$ and $\mathcal{R}_{L,2}(P_{\mathbf{X}',\mathbf{Y}',\mathbf{Y}})$ are $(t\epsilon)$-close. Since $\mathcal{R}_{L,2}(P_{\mathbf{X}',\mathbf{Y},\mathbf{Y}'}) = \mathcal{R}_{L,2}(P_{\mathbf{X}',\mathbf{Y}',\mathbf{Y}})$, $\mathcal{R}_L(P_{\mathbf{X}',\mathbf{Y}'})$ and $\mathcal{R}_L(P_{\mathbf{X}',\mathbf{Y}})$ are $(2t\epsilon)$-close.
In conclusion, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ and $\mathcal{R}_L(P_{\mathbf{X}',\mathbf{Y}'})$ are $((2t+1)\epsilon)$-close. Since the marginal distribution of $P_{\mathbf{X},\mathbf{X}',\mathbf{Y},\mathbf{Y}'}$ on $(\mathbf{X}', \mathbf{Y}')$ is $P_{\mathbf{X}',\mathbf{Y}'}$, which is equal to $Q_{\mathbf{X},\mathbf{Y}}$, by Lemma 1.1, $\mathcal{R}_L(P_{\mathbf{X}',\mathbf{Y}'}) = \mathcal{R}_L(Q_{\mathbf{X},\mathbf{Y}})$, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ and $\mathcal{R}_L(Q_{\mathbf{X},\mathbf{Y}})$ are $(2t+1)\epsilon$-close. This shows that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is continuous in $P_{\mathbf{X},\mathbf{Y}}$. □

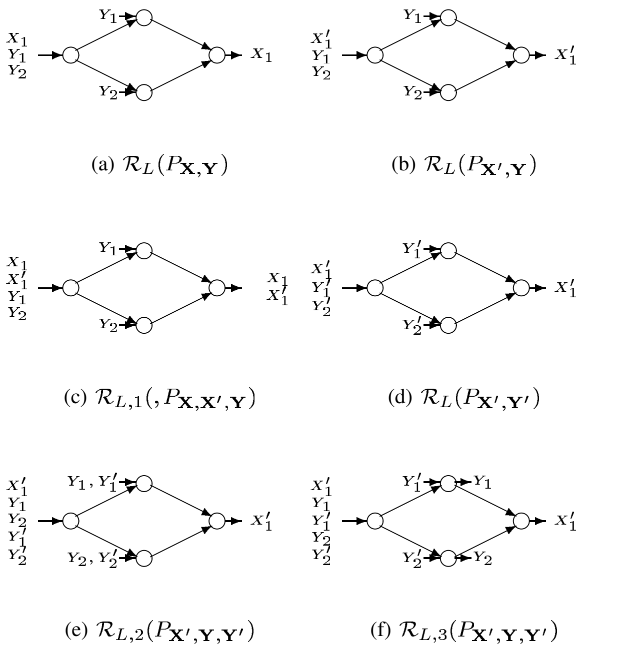
As an easier illustration of the proof of Theorem 3.7, we list the lossless rate regions used in the proof for an example of diamond network in Fig. 2.

We next extend this result to show that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is continuous when $\mathcal{N}$ is a super-source network. First we define $\mathcal{N}_0$ for any network source coding problem $\mathcal{N}$. Note that this definition works for general $\mathcal{N}$, i.e., $\mathcal{N}$ need not be a super-source network. This setting is also useful in the study of outer semi-continuity for general $\mathcal{N}$ in Section III-D.

For any network source coding problem $\mathcal{N} = ((\mathcal{V}, \mathcal{E}), \mathcal{S}, \mathcal{D})$, we consider a new network source coding problem $\mathcal{N}_0 := (G_0, \mathcal{S}_0, \mathcal{D}_0)$ defined as follows, where $G_0 = (\mathcal{V}_0, \mathcal{E}_0)$. First define

$$\mathcal{V}_S := \{v \in \mathcal{E} \mid (v, i) \in \mathcal{S} \text{ for some } 1 \leq i \leq s\}, \tag{3}$$

the set of vertexes in $\mathcal{N}$ each of which has access to some source X_i $(1 \leq i \leq s)$. The vertex set $\mathcal{V}_0$ of $\mathcal{N}_0$ contains $\mathcal{V}$ as a subset and has an extra super-source node v^*. For each $v \in \mathcal{V}_S$, there corresponds a unique intermediate node $v_E(v)$



(a) $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ (b) $\mathcal{R}_L(P_{\mathbf{X}',\mathbf{Y}})$

(c) $\mathcal{R}_{L,1}(,P_{\mathbf{X},\mathbf{X}',\mathbf{Y}})$ (d) $\mathcal{R}_L(P_{\mathbf{X}',\mathbf{Y}'})$

(e) $\mathcal{R}_{L,2}(P_{\mathbf{X}',\mathbf{Y},\mathbf{Y}'})$ (f) $\mathcal{R}_{L,3}(P_{\mathbf{X}',\mathbf{Y},\mathbf{Y}'})$

Fig. 2. The diamond network with different random variable availabilities and demands

in between v^* and v. Thus, the vertex set $\mathcal{V}_0$ is defined as

$$\mathcal{V}_0 := \mathcal{V} \cup \{v^*\} \cup \{v_E(v) \mid v \in \mathcal{V}_S\}.$$

The edge set $\mathcal{E}_0$ contains $\mathcal{E}$ as a subset and has extra links $(v^*, v_E(v))$ and $(v_E(v), v)$ for all $v \in \mathcal{V}_S$, i.e.,

$$\mathcal{E}_0 := \mathcal{E} \cup \{(v^*, v_E(v)), (v_E(v), v) \mid v \in \mathcal{V}_S\}.$$

The super-source node v^* can fully observe $(\mathbf{X}, \mathbf{Y})$, i.e., $(v^*, X_i) \in \mathcal{S}_0$ for all $i \in \{1, \ldots, s\}$ and $(v^*, Y_j) \in \mathcal{S}$ for all $j \in \{1, \ldots, t\}$. For each $v \neq v^*$, $(v, X_i) \notin \mathcal{S}_0$ for all $i \in \{1, \ldots, s\}$, but for all $j \in \{1, \ldots, t\}$, $(v, Y_j) \in \mathcal{S}_0$ if and only if $(v, Y_j) \in \mathcal{S}$. That is,

$$\begin{aligned}\mathcal{S}_0 \;\; := \;\; & \mathcal{S} \cup \{(v^*, X_i) \mid 1 \le i \le s\} \cup \\ & \{(v^*, Y_j) \mid 1 \le j \le t\} - \{(v, X_i) \mid (v, X_i) \in \mathcal{S}\}.\end{aligned}$$

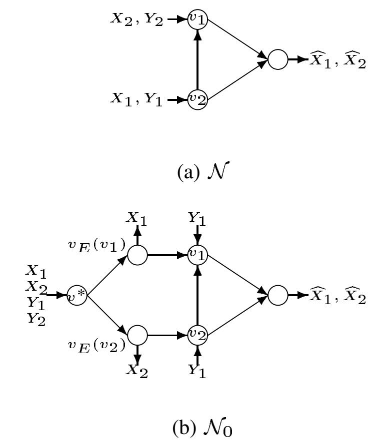
The demand set $\mathcal{D}_0$ of $\mathcal{N}_0$ contains $\mathcal{D}$ as a subset and is defined as

$$\mathcal{D}_0 := \mathcal{D} \cup \{(v_E(v), X_i) \mid (v, X_i) \in \mathcal{S}\}.$$

$\mathcal{N}_0$ is a single-super-source network by definition. An example of turning $\mathcal{N}$ into $\mathcal{N}_0$ is shown in Fig 3. Let $\mathcal{R}_{L,0}(P_{\mathbf{X},\mathbf{Y}})$ denote the lossless rate region for $\mathcal{N}_0$. The network $\mathcal{N}_0$ turns each source accessibility $(v, X_i) \in \mathcal{D}$ at nodes $v \in \mathcal{V}$ in $\mathcal{N}$ other than v^* to demands $(v, X_i) \in \mathcal{D}_0$.

By Theorem 3.7, $\mathcal{R}_{L,0}(P_{\mathbf{X},\mathbf{Y}})$ is continuous with respect to $P_{\mathbf{X},\mathbf{Y}}$. Hence for any sequence $P^{(l)}_{\mathbf{X},\mathbf{Y}}$ converging to $P_{\mathbf{X},\mathbf{Y}}$,

$$\lim_{l\to\infty} \mathcal{R}_{L,0}(P^{(l)}_{\mathbf{X},\mathbf{Y}}) = \mathcal{R}_{L,0}(P_{\mathbf{X},\mathbf{Y}}) \tag{4}$$



(a) $\mathcal{N}$

(b) $\mathcal{N}_0$

Fig. 3. An example of $\mathcal{N}$ and the corresponding $\mathcal{N}_0$

We next define an injection $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ into $\mathcal{R}_{L,0}(P_{\mathbf{X},\mathbf{Y}}) \cap \Upsilon$ for a plane Υ that is defined in (6) and then apply (4) to conclude that the space $\mathcal{R}_{L,0}(P_{\mathbf{X},\mathbf{Y}}) \cap \Upsilon$ is outer semi-continuous in $P_{\mathbf{X},\mathbf{Y}}$.

For $v \in \mathcal{V}_S$, define $\mathbf{X}(v) := (X_i)_{i:(v,X_i)\in\mathcal{S}}$, the source vector available at v in $\mathcal{N}$ and define

$$et(v) = (v^*, v_E(v)), ed(v) = (v_E(v), v).$$

By forcing $R_{et(v)} = R_{ed(v)} = H(\mathbf{X}(v))$ for all $v \in \mathcal{V}_S$ and applying the smoothness of the shape of $\mathcal{R}_L(\mathcal{N}_0, P_{\mathbf{X},\mathbf{Y}})$ (Lemma 2.1), we get

$$\begin{aligned}\limsup_{l\to\infty} & \left(\mathcal{R}_L(\mathcal{N}_0, P^{(l)}_{\mathbf{X},\mathbf{Y}}) \cap \Upsilon\right) \\ & \subseteq \mathcal{R}_L(\mathcal{N}_0, P_{\mathbf{X},\mathbf{Y}}) \cap \Upsilon,\end{aligned} \tag{5}$$

where

$$\Upsilon := \left\{ (R_e)_{e\in\mathcal{E}_0} \mid \begin{array}{l} R_{(v^*,v_E(v))} = R_{(v_E(v),v)} \\ = H(\mathbf{X}(v)) \;\forall\; v \in \mathcal{V}_S \end{array} \right\}. \tag{6}$$

We next connect $\mathcal{R}_L(\mathcal{N}_0, P_{\mathbf{X},\mathbf{Y}})$ to $\mathcal{R}_L(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}})$. By definition of $\mathcal{N}_0$, there is an embedding

$$\varphi : \mathcal{R}_L(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}}) \to \mathcal{R}_L(\mathcal{N}_0, P_{\mathbf{X},\mathbf{Y}}) \cap \Upsilon$$

defined for every $\mathbf{R} = (R_e)_{e\in\mathcal{E}} \in \mathcal{R}_L(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}})$ by $\varphi(\mathbf{R}) := (\varphi_e(\mathbf{R}))_{e\in\mathcal{E}_0}$, where

$$\varphi_e(\mathbf{R}) = \left\{ \begin{array}{ll} R_e & \text{, if } e \in \mathcal{E} \\ H(\mathbf{X}(v)) & \text{, if } e = (v^*, v_E(v)) \\ & \;\;\text{for some } v \in \mathcal{V}_S \\ H(\mathbf{X}(v)) & \text{, if } e = (v_E(v), v) \\ & \;\;\text{for some } v \in \mathcal{V}_S \end{array} \right. .$$

Now when $\mathcal{N}$ is a super-source network, we show that φ is onto, and hence conclude that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is outer semi-continuous in $P_{\mathbf{X},\mathbf{Y}}$. Since inner semi-continuity with respect to $P_{\mathbf{X},\mathbf{Y}}$ is know and proved in Section III-A, we conclude from the following lemma that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is continuous in $P_{\mathbf{X},\mathbf{Y}}$ for any super-source network $\mathcal{N}$.

Theorem 3.8: Let $\mathcal{N}$ be a super-source network. Then

$$\mathcal{R}_{L,0}(P_{\mathbf{X},\mathbf{Y}}) \cap \Upsilon = \varphi(\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})). \tag{7}$$

Proof. For any $\mathbf{R} = (R_e)_{e\in\mathcal{E}_0} \in \mathcal{R}(\mathcal{N}_0, P_{\mathbf{X},\mathbf{Y}}) \cap \Upsilon$. We want to show that the rate vector $(R_e)_{e\in\mathcal{E}}$ is in $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$. For any $\epsilon > 0$, let $\mathcal{C}_n$ be a length-n rate-($\mathbf{R} + \epsilon \cdot \mathbf{1}$) block code such that the error probability of successful reconstructions is less than ϵ/m. For every $v \in \mathcal{V}_S$, let $F^{(n)}_{ed(v)}$ be the encoded message on the link $ed(v)$. Since $\mathbf{R} \in \Upsilon$,

$$H(F^{(n)}_{ed(v)}) \le n(H(\mathbf{X}(v)) + \epsilon).$$

On the other hand, by Fano's inequality,

$$H(\mathbf{X}(v)^n | F^{(n)}_{ed(v)}) < n\epsilon.$$

Hence

$$\begin{aligned} &H(F^{(n)}_{ed(v)} | \mathbf{X}(v)^n) \\ &\quad = H(F^{(n)}_{ed(v)}, \mathbf{X}(v)^n) - H(\mathbf{X}(v)^n) \\ &\quad = H(F^{(n)}_{ed(v)}) + H(\mathbf{X}(v)^n | F^{(n)}_{ed(v)}) - H(\mathbf{X}(v)^n) \\ &\quad \le n\left(H(\mathbf{X}(v)) + \epsilon + \epsilon - H(\mathbf{X}(v))\right) \\ &\quad = 2n\epsilon. \end{aligned}$$

Therefore, there exists a function of $(\mathbf{X}^n, \mathbf{Y}^n)$, denoted by U_n, such that

$$\begin{aligned} &H(F^{(n)}_{ed(v)} | \mathbf{X}(v)^n, U_n) = 0 \\ &H(U_n) < 2n\epsilon. \end{aligned}$$

It means that for every $v \in \mathcal{V}_S$, by transmitting additional random variable U_n on the link (v^*, v), the messages transmitted over $et(v)$ and $ed(v)$ can be functions of $\mathbf{X}(v)^n$ only such that the node v can still get $F^{(n)}_{ed(v)}$ with sufficiently high probability. (See Fig. 4.) Since $H(U_n) < 2n\epsilon$, the additional rate for transmitting U_n can be made arbitrarily small, and so the modified coding scheme has the same rate $\mathbf{R}$. Therefore, $(R_e)_{e\in\mathcal{E}} \in \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$. □

From Theorem 3.8, we conclude this subsection by Theorem 3.9.

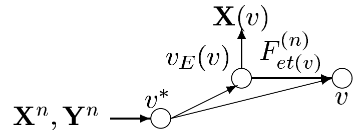
Theorem 3.9: . Let $\mathcal{N}$ be a super-source network. Then for any sequence of distributions $\{P^{(l)}_{\mathbf{X},\mathbf{Y}}\}_{l=1}^{\infty} \subset \mathcal{M}$ that converges to $P_{\mathbf{X},\mathbf{Y}}$,

$$\lim_{l\to\infty} \mathcal{R}_L(P^{(l)}_{\mathbf{X},\mathbf{Y}}) = \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$$

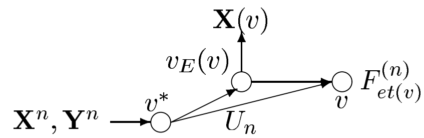
Proof. By Theorem 3.8 and equation (5), $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is outer semi-continuous in $P_{\mathbf{X},\mathbf{Y}}$. Since $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is inner semi-continuous in $P_{\mathbf{X},\mathbf{Y}}$ by Theorem 3.4, we conclude that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is continuous in $P_{\mathbf{X},\mathbf{Y}}$. □

C. The Vanishment Conjecture

In Section III-B, we prove that the injection φ is onto for super-source network, which leads to $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is continuous in $P_{\mathbf{X},\mathbf{Y}}$ by Theorem 3.9. Note that he only difference between super-source networks and general network source



(a) Original Coding Strategy



(b) Modified Coding Strategy with the same rate

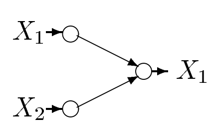
Fig. 4. The descending process

coding problems is that there are direct links from super source node v^* to all the nodes v in $\mathcal{V}_S$. (See definition of $\mathcal{V}_S$ in (3)). In any super-source network, if we force the rates on those direct links (v^*, v) $(v \in \mathcal{V}_S)$ to be negligible, then the resulting rate regions are supersets of the rate regions corresponding to the network without those direct links (v^*, v) $(v \in \mathcal{V}_S)$. The Vanishment Conjecture states that those links of negligible rates can be removed. We formalize this idea in the following paragraph.

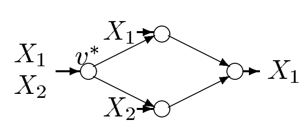
Let $\mathcal{N}$ be a network source coding problem. Consider the super-source network $\mathcal{N}'$ defined from $\mathcal{N}$ as follows

$$\begin{aligned} &\mathcal{N}' = ((\mathcal{V}', \mathcal{E}'), \mathcal{S}', \mathcal{D}') \\ &\mathcal{V}' = \mathcal{V} \cup \{v^*\} \\ &\mathcal{E}' = \mathcal{E} \cup \{(v^*, v) \mid v \in \mathcal{V}_S\} \\ &\mathcal{S}' = \mathcal{S} \cup \{(v^*, X_i) \mid i \in \{1, \ldots, s\}\} \\ &\qquad \cup \{(v^*, Y_j) \mid j \in \{1, \ldots, s\}\} \\ &\mathcal{D}' = \mathcal{D}. \end{aligned}$$

By above definition, $\mathcal{N}'$ is the super-source network modified from $\mathcal{N}$ by adding the super source node v^* and the links (v^*, v) for all $v \in \mathcal{V}_S$. Fig. 5 shows an example of $\mathcal{N}'$ when $\mathcal{N}$ is the two-terminal network source coding problem.



(a) Two-terminal Network



(b) Diamond Network

Fig. 5. The two-terminal problem and diamond networks.

Let $\mathcal{R}'_L(P_{\mathbf{X},\mathbf{Y}})$ denote the lossless rate region for $\mathcal{N}'$. For any $\mathbf{R} = (R_e)_{e\in\mathcal{E}}$, define the rate vector $\zeta(\mathbf{R}) =$

$(\zeta_e(\mathbf{R}))_{e\in\mathcal{E}'}$ as

$$\zeta_e(\mathbf{R}) := \begin{cases} R_e & \text{, if } e \in \mathcal{E} \\ 0 & \text{, if } e \notin \mathcal{E} \end{cases}.$$

Define the plane

$$\Delta := \{R_{(v^*,v)} = 0 \mid v \in \mathcal{V}_S\}. \tag{8}$$

Then the map of ζ restricted on $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$

$$\zeta : \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}}) \to \mathcal{R}'_L(P_{\mathbf{X},\mathbf{Y}}) \cap \Delta$$

is an injection. Now define

$$\zeta(\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})) := \{\zeta(\mathbf{R}) \mid \mathbf{R} \in \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})\}. \tag{9}$$

By definition, $\zeta(\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})) \subseteq \mathcal{R}'_L(P_{\mathbf{X},\mathbf{Y}}) \cap \Delta$.

We conjecture that this map is onto as well. See the following definition.

Definition 3.10: The Vanishment Conjecture states that

$$\zeta(\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})) = \mathcal{R}'_L(P_{\mathbf{X},\mathbf{Y}}) \cap \Delta. \tag{10}$$

D. Outer Semi-Continuity

We investigate outer semi-continuity here. The main idea is to rely on the result in Section III-B and the Vanishment Conjecture. Similar to the approach in Section III-B, we begin by first defining the corresponding super-source network $\mathcal{N}'$ of network source coding problem $\mathcal{N}$ and embed $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ into $\mathcal{R}'_L(P_{\mathbf{X},\mathbf{Y}}) \cap \Delta$ by using ζ defined in (9), where Δ is the plane defined in (8). Now since $\mathcal{N}'$ is a super-source network, by Theorem 3.9, $\mathcal{R}'_L(P_{\mathbf{X},\mathbf{Y}})$ is continuous in $P_{\mathbf{X},\mathbf{Y}}$, which leads to

$$\limsup_{l\to\infty} \mathcal{R}'_L(P^{(l)}_{\mathbf{X},\mathbf{Y}}) \cap \Delta \subseteq \mathcal{R}'_L(P_{\mathbf{X},\mathbf{Y}}) \cap \Delta \tag{11}$$

for every sequence $\{P^{(l)}_{\mathbf{X},\mathbf{Y}}\}_{l=1}^{\infty} \subset \mathcal{M}$ that converges to $P_{\mathbf{X},\mathbf{Y}}$. We then have the following theorem.

Theorem 3.11: Let $\mathcal{N}$ be a non-functional network source coding problem. Let $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$. Let $\{P^{(l)}_{\mathbf{X},\mathbf{Y}}\} \subset \mathcal{M}$ be a sequence that converges to $P_{\mathbf{X},\mathbf{Y}}$. Then the Vanishment Conjecture implies that

$$\limsup_{l\to\infty} \mathcal{R}_L(P^{(l)}_{\mathbf{X},\mathbf{Y}}) \subseteq \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}}).$$

Proof. It is a direct consequence of equations (11) and (10). □

Theorem 3.12: Let $\mathcal{N}$ be a non-functional network source coding problem. Let $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$. Let $\{P^{(l)}_{\mathbf{X},\mathbf{Y}}\} \subset \mathcal{M}$ be a sequence that converges to $P_{\mathbf{X},\mathbf{Y}}$. Then the Vanishment Conjecture implies that

$$\lim_{l\to\infty} \mathcal{R}_L(P^{(l)}_{\mathbf{X},\mathbf{Y}}) = \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}}).$$

Proof. It is a direct consequence of Theorems 3.4 and 3.11. □

APPENDIX

A. Lemmas for Theorem 3.7

In order to prove Theorem 3.7, the first step is to show that when two distributions $P_{\mathbf{X},\mathbf{Y}}$ and $Q_{\mathbf{X},\mathbf{Y}}$ are sufficiently close, there exists a joint distribution $P_{\mathbf{X},\mathbf{X}',\mathbf{Y},\mathbf{Y}'}$ such that the two pairs of random variables $(\mathbf{X},\mathbf{Y})$ and $(\mathbf{X}',\mathbf{Y}')$ have distributions $P_{\mathbf{X},\mathbf{Y}}$ and $Q_{\mathbf{X},\mathbf{Y}}$, respectively, and satisfy that they are equal to each other componentwise with high probability.

Lemma 1.1: Let $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$. For any $\epsilon > 0$, there exists a $\delta > 0$ such that for any $Q_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$ satisfying $||P_{\mathbf{X},\mathbf{Y}} - Q_{\mathbf{X},\mathbf{Y}}|| < \delta$, there exists a joint distribution $P_{\mathbf{X},\mathbf{X}',\mathbf{Y},\mathbf{Y}'}$ satisfying the following properties:

1) *For every $\mathbf{x} \in \{1,\ldots,m\}^s$ and every $\mathbf{y} \in \{1,\ldots,m\}^t$,*
$$\sum_{\mathbf{x}',\mathbf{y}'} P_{\mathbf{X},\mathbf{X}',\mathbf{Y},\mathbf{Y}'}(\mathbf{x},\mathbf{x}',\mathbf{y},\mathbf{y}') = P_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}).$$
2) *For every $\mathbf{x}' \in \{1,\ldots,m\}^s$ and every $\mathbf{y}' \in \{1,\ldots,m\}^t$,*
$$\sum_{\mathbf{x},\mathbf{y}} P_{\mathbf{X},\mathbf{X}',\mathbf{Y},\mathbf{Y}'}(\mathbf{x},\mathbf{x}',\mathbf{y},\mathbf{y}') = Q_{\mathbf{X},\mathbf{Y}}(\mathbf{x}',\mathbf{y}').$$
3) $\Pr(X_i = X'_i) > 1-\epsilon \;\forall\, i \in \{1,\ldots,s\}$.
4) $\Pr(Y_j = Y'_j) > 1-\epsilon \;\forall\, j \in \{1,\ldots,t\}$.

Proof. We demonstrate the result by finding a conditional distribution $P_{(\mathbf{X}',\mathbf{Y}')|(\mathbf{X},\mathbf{Y})}$ for which the joint distribution

$$P_{\mathbf{X},\mathbf{X}',\mathbf{Y},\mathbf{Y}'} = P_{\mathbf{X},\mathbf{Y}} P_{(\mathbf{X}',\mathbf{Y}')|(\mathbf{X},\mathbf{Y})}$$

satisfies the desired properties. Rename the alphabet of $(\mathbf{X},\mathbf{Y})$ as $\{a_1,a_2,\ldots,a_l\}$, where $l = m^2$. Let $p_r = P_{\mathbf{X},\mathbf{Y}}(a_r)$ and $q_r = Q_{\mathbf{X},\mathbf{Y}}(a_r)$ for all $r \in \{1,\ldots,l\}$. Enumerate the alphabets of $(\mathbf{X},\mathbf{Y})$ such that there exist $1 \le l_1 \le l_2 \le l$ such that

$$\begin{aligned} p_r &< q_r \;\forall\, r \in \{1,\ldots,l_1\} \\ p_r &= q_r \;\forall\, r \in \{l_1+1,\ldots,l_2\} \\ p_r &> q_r \;\forall\, r \in \{l_2+1,\ldots,l\}. \end{aligned}$$

Define the conditional distribution $P_{(\mathbf{X}',\mathbf{Y}')|(\mathbf{X},\mathbf{Y})}$ as follows

$$\begin{aligned} &P_{(\mathbf{X}',\mathbf{Y}')|(\mathbf{X},\mathbf{Y})}(a_r|a_r) = 1 \;\forall\, r \in \{1,\ldots,l_2\} \\ &P_{(\mathbf{X}',\mathbf{Y}')|(\mathbf{X},\mathbf{Y})}(a_r|a_r) = \frac{q_r}{p_r} \; r \in \{l_2+1,\ldots,l\} \\ &P_{(\mathbf{X}',\mathbf{Y}')|(\mathbf{X},\mathbf{Y})}(a_u|a_r) = \frac{q_u - p_u}{\sum_{u'=1}^{l_1}(q_{u'} - p_{u'})}(1 - \frac{q_r}{p_r}) \\ &\forall\, r \in \{l_2+1,\ldots,l\} \;\forall\, u \in \{1,\ldots,l_1\}. \end{aligned}$$

Let $(\mathbf{X}',\mathbf{Y}')$ be another random vector such that the joint distribution of $(\mathbf{X},\mathbf{X}',\mathbf{Y},\mathbf{Y}')$ is defined by

$$\begin{aligned} &P_{\mathbf{X},\mathbf{X}',\mathbf{Y},\mathbf{Y}'}(\mathbf{x},\mathbf{x}',\mathbf{y},\mathbf{y}') \\ &\quad := \; P_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) P_{(\mathbf{X}',\mathbf{Y}')|(\mathbf{X},\mathbf{Y})}(\mathbf{x}',\mathbf{y}'|\mathbf{x},\mathbf{y}) \\ &\quad\quad\; \forall \mathbf{x},\mathbf{y},\mathbf{x}',\mathbf{y}'. \end{aligned}$$

Then

$$\Pr\{(\mathbf{X},\mathbf{Y}) = (\mathbf{X}',\mathbf{Y}')\} \;=\; 1 - \sum_{r=l_2+1}^{l} (p_r - q_r).$$

Now let $\delta > 0$ be sufficiently small such that

$$\sum_{r=l_2+1}^{l} (p_r - q_r) < \epsilon$$

whenever $||P_{\mathbf{X},\mathbf{Y}} - Q_{\mathbf{X},\mathbf{Y}}|| < \delta$. Then

$$\Pr\{X_i = X'_i\} \geq \Pr\{\mathbf{X} = \mathbf{X}', \mathbf{Y} = \mathbf{Y}'\} > 1 - \epsilon$$

for all $i \in \{1, \ldots, s\}$. Similarly,

$$\Pr\{Y_j = Y'_j\} \geq \Pr\{\mathbf{X} = \mathbf{X}', \mathbf{Y} = \mathbf{Y}'\} > 1 - \epsilon$$

for all $j \in \{1, \ldots, t\}$. Finally, for all $\mathbf{x}' \in \{1, \ldots, m\}^s$ and for all $\mathbf{y}' \in \{1, \ldots, m\}^t$

$$\sum_{\mathbf{x},\mathbf{y}} P_{\mathbf{X},\mathbf{X}',\mathbf{Y},\mathbf{Y}'}(\mathbf{x}, \mathbf{x}', \mathbf{y}, \mathbf{y}') = Q_{\mathbf{X},\mathbf{Y}}(\mathbf{x}', \mathbf{y}').$$

□

We state without proving the following lemma that shows that when a side information Y_t appears only at the super source node v^* in a single-super-source network, Y_t doesn't improve the whole lossless rate region.

Lemma 1.2: Let $\mathcal{N}$ be a single-super-source network source coding problem as in Theorem 3.7. Suppose that $(v^*, Y_t) \in \mathcal{S}$ but for all $v \neq v^*$, $(v, Y_t) \notin \mathcal{S}$. Then the network source coding problem that has s sources and $t-1$ side-information random variables defined as

$$\mathcal{N}' := (G, \mathcal{S} - \{(v^*, Y_t)\}, \mathcal{D})$$

satisfies that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}}) = \mathcal{R}'_L(P_{\mathbf{X},(Y_1,\ldots,Y_{t-1})})$ for all $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$, where $P_{\mathbf{X},(Y_1,\ldots,Y_{t-1})}$ is the marginal distribution of $P_{\mathbf{X},\mathbf{Y}}$ on $(\mathbf{X}, (Y_1, \ldots, Y_{t-1}))$ and $\mathcal{R}'_L(P_{\mathbf{X},(Y_1,\ldots,Y_{t-1})})$ is the lossless rate region for the network source coding problem $\mathcal{N}'$.

The remainder of Appendix A is the following lemma, which demonstrates some relations between the rate regions for the networks $\mathcal{N}$, $\mathcal{N}_1$, and $\mathcal{N}_3$ defined in the proof of Theorem 3.7. That result leads to conclude the proof of Theorem 3.7. Since the results are straightforward, we state the Lemma without proving it.

Lemma 1.3: Let $\mathbf{X}$, $\mathbf{X}'$, $\mathbf{Y}$, $\mathbf{Y}'$, $\mathcal{N}$, $\mathcal{N}_1$, $\mathcal{N}_3$, $\mathbf{D}^{(1)}$, and $\mathbf{D}^{(3)}$ be defined as in the proof of Theorem 3.7. Let $\mathcal{R}$, $\mathcal{R}_1$, and $\mathcal{R}_3$ denote the rate regions for networks $\mathcal{N}_1$, $\mathcal{N}_2$, and $\mathcal{N}_3$, respectively.

(a)

$$\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}}) \subseteq \mathcal{R}_{L,1}(P_{\mathbf{X},\mathbf{X}',\mathbf{Y}}).$$

(b) For $\epsilon > 0$ sufficiently small,

$$\mathcal{R}_{L,1}(P_{\mathbf{X},\mathbf{X}',\mathbf{Y}}) + \epsilon \cdot \mathbf{1} \subseteq \mathcal{R}_L(P_{\mathbf{X}',\mathbf{Y}}).$$

(c)

$$\mathcal{R}_{L,3}(P_{\mathbf{X}',\mathbf{Y},\mathbf{Y}'}) \subseteq \mathcal{R}_L(P_{\mathbf{X}',\mathbf{Y}'}).$$

B. Lemma for Section III-D

We describe and prove a lemma useful in Section III-D. In this lemma, we prove the smoothness of boundary surfaces for $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$.

Let $\mathcal{N}$ be a non-functional network source coding problem. For every $\widehat{e} \in \mathcal{E}$, define the vector $\jmath(\widehat{e}, r) = (\jmath(\widehat{e}, r)_e)_{e \in \mathcal{E}}$ for $r > 0$ as follows

$$\jmath(\widehat{e}, r)_e = \begin{cases} 0, & \text{if } e \neq \widehat{e} \\ r, & \text{if } e = \widehat{e} \end{cases}.$$

For any $r \geq r_0$, define

$$\begin{aligned} \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})[r, \widehat{e}] \;\; := \;\; & \{\mathbf{R} = (R_e)_{e \in \mathcal{E} - \{\widehat{e}\}} \mid \mathbf{R}_{\widehat{e}} = 0, \\ & \mathbf{R} + \jmath(\widehat{e}, r) \in \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})\} \end{aligned}$$

The following lemma shows the smoothness of the boundary surfaces of $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$.

Lemma 2.1: Let $\mathcal{N}$ be a network source coding problem and $\widehat{e} \in \mathcal{E}$ be a fixed edge. Suppose $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$. Let $r_0 \geq 0$ be a nonnegative real number such that

$$\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}}) \cap \{R_{\widehat{e}} = r_0\} \neq \emptyset.$$

Then $\lim_{r \to r_0} \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})[r, \widehat{e}] = \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})[r_0, \widehat{e}]$.

Proof. For simplicity, let $\mathcal{R}[r]$ denote $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})[r, \widehat{e}]$. By definition, $\mathcal{R}[r_0] \subseteq \mathcal{R}[r]$ for all $r \geq r_0$. Therefore, $\mathcal{R}[r_0] \subseteq \liminf_{r \to r_0} \mathcal{R}[r]$. On the other hand, since $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is closed, $\limsup_{r \to r_0} \mathcal{R}[r] \subseteq \mathcal{R}[r_0]$. In conclusion, $\lim_{r \to r_0} \mathcal{R}[r] = \mathcal{R}[r_0]$. This completes the proof. □

References

[1] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. In *IRE National Convention Record*, volume 4, pages 142–163, 1959.

[2] W. Gu and M. Effros. On the continuity of achievable rate regions for source coding over networks. In *Proceedings of the IEEE Information Theory Workshop*, Tahoe, CA, September 2007.

[3] T. Han and K. Kobayashi. A dichotomy of functions $f(x, y)$ of correlated sources (x, y). *IEEE Transactions on Information Theory*, IT-33(1):69–76, January 1987.

[4] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, IT-19:471–480, 1973.

[5] R. Ahlswede, N. Cai, S. Y. R. Li, and R. Yeung. Network information flow. *IEEE Transactions on Information Theory*, IT-46(4):1204–1216, July 2000.

[6] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong. A random linear network coding approach to multicast. *IEEE Transactions on Information Theory*, IT-52(10):4413–4430, October 2006.

[7] M. Bakshi and M. Effros. On achievable rates for multicast in the presence of side information. In *Proceedings of the IEEE International Symposium on Information Theory*, Toronto, Canada, July 2008.

[8] R. Ahlswede and J. Körner. Source coding with side information and a converse for degraded broadcast channels. *IEEE Transactions on Information Theory*, IT-21(6):629–637, November 1975.

[9] R. M. Gray and A. D. Wyner. Source coding for a simple network. *Bell System Technical Journal*, 53(9):1681–1721, November 1974.

[10] A. Sgarro. Source coding with side information at several decoders. *IEEE Transactions on Information Theory*, IT-23(2):179–182, March 1977.

[11] A. Orlitsky and J. R. Roche. Coding for computing. *IEEE Transactions on Information Theory*, IT-47(3):903–917, March 2001.

[12] J. Chen and A. B. Wagner. A continuity theorem and its applications to gaussian newtork source coding problems. In *Proceedings of the Information Theory and Applications*, San Diego, California, February 2008.

[13] J. Chen and A. B. Wagner. A semicontinuity theorem and its application to network source coding. In *Proceedings of the IEEE International Symposium on Information Theory*, Toronto, Canada, July 2008.